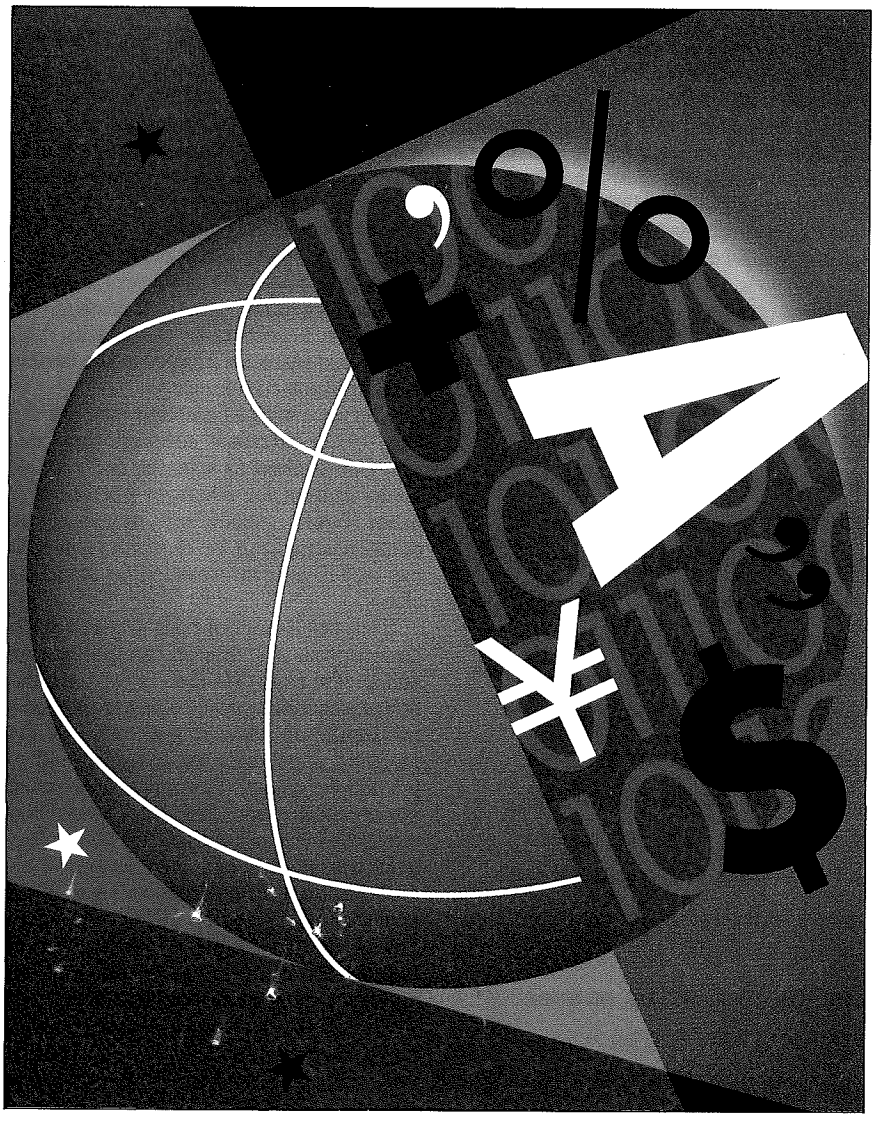
# BROWSING THROUGH TERABYTES

## Wide-area information servers open a new frontier in personal and corporate information services

RICHARD MARLON STEIN

The Library of Congress archives roughly 25 terabytes in its collection. To browse through this volume on your own would be nearly impossible. Wide-area information servers supply the means to achieve this goal by providing the user-interface structure and underlying information-retrieval protocol necessary to automatically collate, collect, and integrate diverse data streams. WAISes can distill the contents of vast archives into neatly manageable and browsable folders.

On-line information services, such as BIX and CompuServe, attest to the need for this kind of technology. Information has acquired a commodity-like status. While not on a par with wheat, pork bellies, or gold futures, the information-service industry fills a vital role. The next phase of information commerce will add WAIS capabilities to existing on-line services, opening a new frontier in personal and corporate information services.



### Intentions and Goals

Initiated in early 1989, the WAIS engineering effort is spearheaded by Thinking Machines (Cambridge, MA), the manufacturer of the Connection Machine, a massively parallel supercomputer (see reference 1). The principal goal of the research project is to demonstrate "how current technology can be used to open a market of information services that will allow a user's workstation to act as librarian and information collection agent from a large number of sources." (See reference 2.) WAISes aim to enhance existing information services and provide a utilitarian mechanism for the industry.

*continued*

ILLUSTRATION: RON CHAN © 1991

Information servers already provide direct access to many databases and archive structures. You can easily check the local weather, make travel reservations, obtain entertainment schedules, or browse through the latest stock-market quotes on-line. These services are highly interactive, charging users on the basis of minutes spent on-line, and each has a unique user interface.

WAISes alleviate unnecessary user interaction through a predominantly computer-to-computer approach to remote information retrieval. By minimizing human interaction with a remote information server, they handle requests for information expeditiously and inexpensively. WAISes also alleviate unnecessary complexity by moving all user interaction to the local workstation and by having WAIS software handle all transactions with the remote server.

On-line servers are limited in their connectivity. While many services, such as BIX, CompuServe, and AppleLink, incorporate wide-area network structures, sharing information between different services is not a wholly transparent option. This restriction constrains information commerce and hampers the circulation of potentially useful ideas.

WAISes circumvent this barrier with a standard information-exchange protocol that offers unlimited connectivity and retrieval functionality. All servers can apply the WAIS protocol to their archive structures to conduct information retrieval. (Unlimited connectivity also raises concerns of security and privacy. See the text box "The Right to Privacy" on page 160.)

Organized and coherent information of topical importance has value. Individuals and companies should be able to market their information to the widest possible audience. Current on-line services can't easily accomplish this, since their connectivity is restricted.

To direct your information to the best marketplace, you could subscribe to multiple on-line sources and post the same message on all of them. But it would be more efficient to post the data on one server and have the data, or an abstract of it, broadcast to the others. Using the WAIS protocol, WAISes facilitate this server function.

Suppose, for example, you have reviewed the latest set of RISC microprocessor benchmarks, taking note of specific architectural advantages, and you wish to make this information available to others. The benchmark review is kept on your home computer (i.e., the local WAIS), which is equipped with WAIS technology. The nearest remote WAIS, a hub within a network of servers, also has a folder for RISC microprocessors. So you make a posting to the nearest hub server that inserts a pointer to the review on your home computer.

Everyone with a computer running the WAIS user-interface software can present information to a server and receive compensation for whatever portion of it other WAIS subscribers access. The compensation can be monetary, or you can barter your information for someone else's.

Even publishers of books, magazines, newspapers, and music can participate and profit from WAISes. For example, how much money could a newspaper save in circulation costs if you received the morning paper electronically instead of printed on paper? Similarly, how much money could a book publisher save if you purchased a new best-selling novel electronically instead of at a bookstore?

Traditional information delivery is expensive, and costs are rising. The U.S. Postal Service frequently raises its fees to cover increases in the cost of handling and transporting information. Traditional information transport also represents a significant fraction of transport volume and collateral energy consumption. Moving information electronically can result in enormous savings.

Computer networks such as Internet are conduits of information transport. To replace manual transportation methods, the existing electronic infrastructure must accommodate the newly anticipated volume of traffic. Plans for "a national network of data superhighways," which will be installed within the next few years, are under way (see references 3 and 4).

A principal motivation for WAIS technology is to be able to retrieve topical information for research or investigation, not just to deliver consumable items like newspapers or books. Toward this end, WAISes rely on a novel structure for information retrieval, the *dynamic folder*.

To use a WAIS, you formulate a question (see figure 1), find the information servers that provide satisfactory responses, and create a dynamic folder. The purpose of the dynamic folder is to constantly or periodically update its contents with new material on the subject.

Formulating a question is natural to us all. The difficult part is locating the pertinent information to answer it. Manually locating the information can be laborious and tedious. WAISes automate the search-and-retrieval process. To determine which servers hold the information most pertinent to your question, and where you should submit dynamic folders, you may want to consult *server directories*.

## Server Directories

WAIS directories are servers that support a directory-services function. They are indexes to other services within the WAIS network and are organized to help you locate information. Like telephone-directory services, WAIS directories list pointers to servers, which are grouped according to content and function.

A *directory-entry header* contains sufficient data to describe the service, such as an English-language description of the server, the parent server (if the server is a subsidiary of a larger one), related servers, contact information (including networks and human-interface points), and cost information.

The local workstation, when equipped with a WAIS, should maintain a directory entry that includes the directory-entry header, a locally determined rank, subscription information (if any), user comments, and the time of last contact. You can use this information to decide whether to contact the server and how to handle the responses.

By using content navigation, you can find the most appropriate server to

**BYTE ACTION SUMMARY**

The next phase of information commerce will add wide-area information server capabilities to existing on-line services. WAISes provide the user-interface structure and the underlying information-retrieval protocol necessary to automatically collate, collect, and integrate information from various sources. When these are implemented, you should be able to directly access such sources as the Library of Congress and the myriad of newspapers, journals, and books.

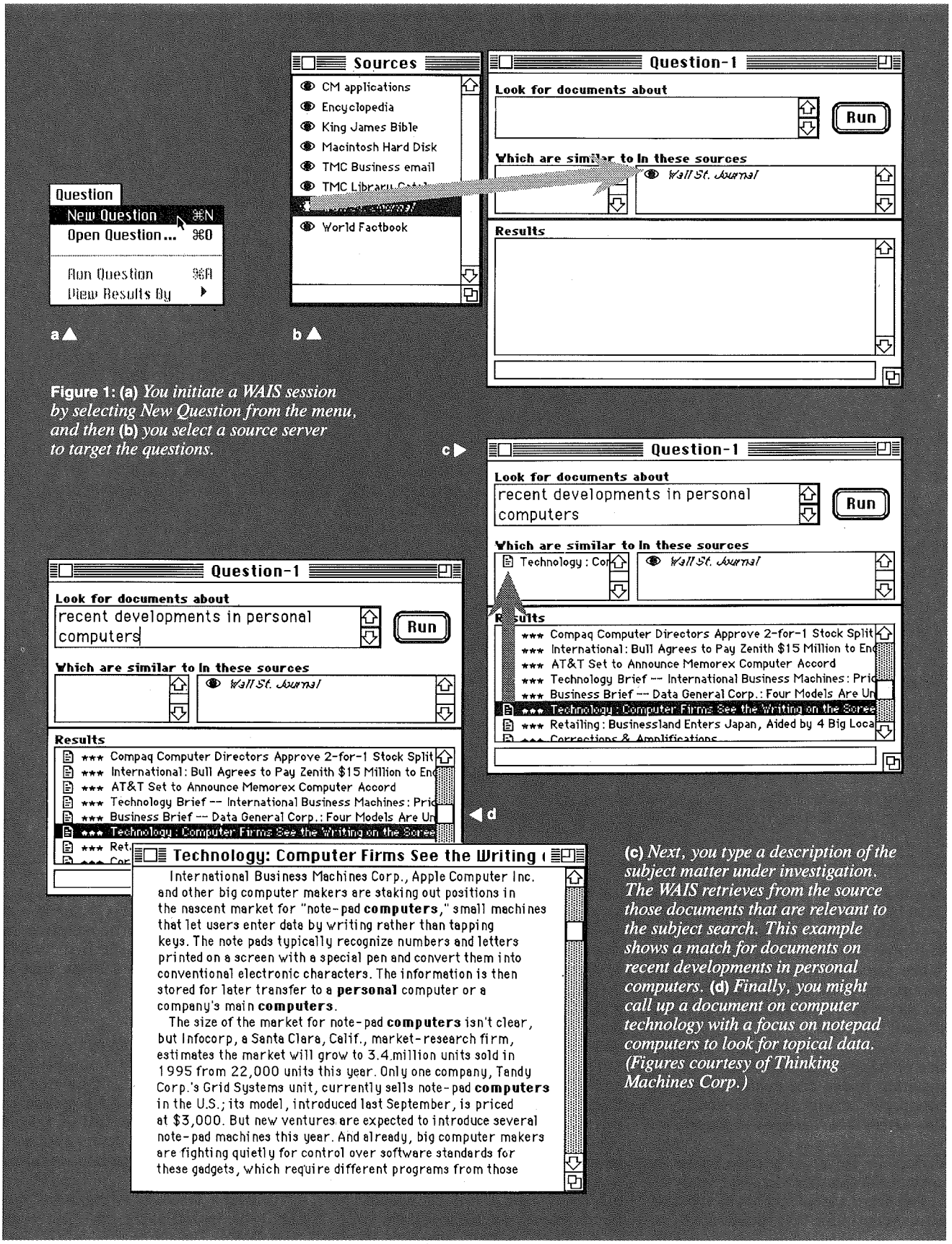



**Figure 1: (a)** *You initiate a WAIS session by selecting New Question from the menu, and then* **(b)** *you select a source server to target the questions.* **(c)** *Next, you type a description of the subject matter under investigation. The WAIS retrieves from the source those documents that are relevant to the subject search. This example shows a match for documents on recent developments in personal computers.* **(d)** *Finally, you might call up a document on computer technology with a focus on notepad computers to look for topical data. (Figures courtesy of Thinking Machines Corp.)*

# The Right to Privacy

WAIStation, a prototype user interface developed by the Thinking Machines wide-area information server project staff, embodies many functional aspects of WAIS technology. Forming and refining queries via relevance feedback, server selection, and dynamic folders are the principal features that this prototype supports. These assets provide a powerful tool set for information retrieval. While WAIStation achieves several desirable technical goals, the security and privacy issues have not yet received serious attention and need refinement.

Security and privacy issues are not specific to WAIStation or WAISes in general, but are endemic, topical concerns of the information-retrieval industry as a whole. WAIS technology seeks to extend connectivity through the WAIS protocol, thus intensifying the urgency of security measures and standards. Greater connectivity promotes information commerce, but it also adds to the risk of compromising the privacy and confidentiality of electronic transactions.

Individuals and corporations that subscribe to WAISes must safeguard proprietary information. The tendency to organize information within a computer for ease of access or to act as a convenient archive creates a security and privacy dilemma. And if the sensitive data is located on a machine with high connectivity, the risk is multiplied.

A WAIStation that holds personal information, such as tax forms, diaries, business transactions, medical records, or bank accounts, must be protected from intrusion by unauthorized individuals. A computer system storing this information "knows" more about you than you can instantly recall. Access to this personal data must be protected, controlled, and limited to authorized individuals.

The WAIS protocol is an application-layer protocol that runs over X.25 communications, modems, or IEEE 802.3 (Ethernet) backbones. Residing beneath this protocol is the WAIStation host computer and operating system. Extracting information from the server depends on access granted through a recognition and authentication system that the host computer operates. Only authorized subscribers can access information from the server.

The WAIS protocol is stateless, so each transaction, whether a query or document-retrieval process, exists in a separate context at the server. Subversion of the WAIS protocol, whether intentional or accidental, might unlock or bypass a server's native file-system protection structure. If it did, the entire archive contents would be available to the intruding party.

The WAIS protocol should be noncorruptible and should detect privileged transactions (i.e., those data streams that possess restricted command sequences). However, to be effective as a noncorruptible application-layer protocol, the underlying computer system must also be unbreachable.

Unfortunately, you cannot always guarantee protection. In 1988, a virus introduced through a known port assaulted computer systems attached to Internet. Subsequent sleuthing discovered that a remote system could activate the debug mode of the Unix mailer, forcing the instigator into a privileged state. The debug mode then permitted the virus to propagate and multiply.

Can a rogue dynamic folder, fashioned after the Internet virus, intentionally access information from strategic servers running WAIS software? How will WAISes safeguard information against illegal intrusion?

The right to privacy is inalienable, and WAIS technology or any enabling system that promotes information commerce must preserve it. A cautionary approach toward implementating WAIS technology is necessary and appropriate. Several legal issues must be addressed to secure both privacy and fair business practice.

handle a query. For example, a question on RISC microprocessor benchmarks would list directory entries for servers as well as pointers to articles on the subject. When you retrieve a document, the directory entry is also provided. Thus, you obtain ranking information for questions of similar content.

Each server, then, contains information of value to certain subscribers. The dynamic folder can continuously poll newspaper servers for new articles as they arrive from the news wires, while it would probably query a dictionary or encyclopedia server only once, since the content changes much less frequently.

Policing the large number of anticipated servers (in the tens of thousands) requires an independent quality-control mechanism. An audit of the server directory would reflect any server that frequently returns erroneous information or does not perform. An independent agency like *Consumer Reports*, the Better Business Bureau, or other watchdog groups could create *rating servers*, which monitor and rate other servers in the directory.

These rating servers resemble movie and TV critics. Consumers acquire confidence in the reports and reviews that certain critics issue because they share similar tastes. Just as moviegoers start to trust a particular reviewer who has agreed with them on past movies, WAIS users will begin to trust the specific rating services that agree with them.

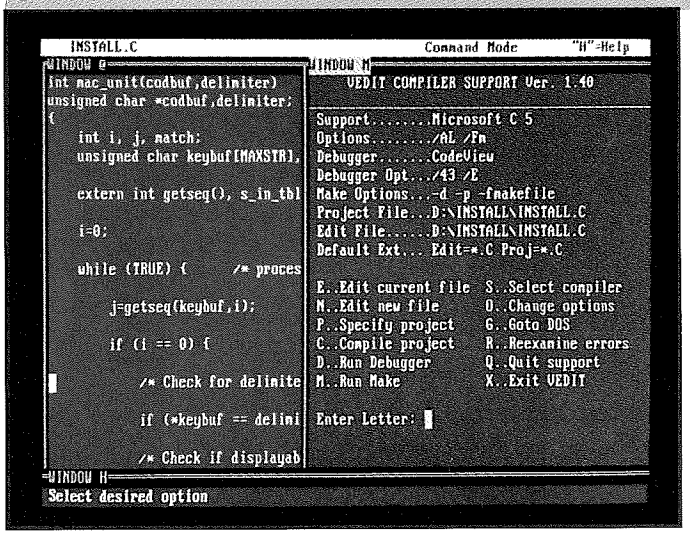
A subscriber base generates income for a server. The rating servers will attract subscribers as well, for they direct trends in the information marketplace. In fact, they may become the first "information speculators" as a by-product of WAIS technology.

### Dynamic Folders

A folder, like those found on the Macintosh, provides the WAIS framework for organizing questions. A folder is a repository for documents. A file system, in the Macintosh sense, is full of folders organized in a tree structure that supports an efficient document-location mechanism.

To find a document within a file system, you typically use the `find` command under Unix or Finder on the Mac.

# *Finally, a fast, powerful text editor that integrates your favorite programming tools and uses no memory!*




The new VEDIT family of text editors offers stunning performance, versatility and ease of use. Completely written in assembly language, they are small and lightning fast. (3 to 30 times faster than other editors on large files where speed really counts.) Edit text and binary files of any size, even 100+ megabytes. Installation is trivial; VEDIT.EXE and an optional help file are all you need - no overlays, no configuration files, no environment variables.

For programmers, the new compiler support in VEDIT and VEDIT PLUS is a breakthrough. Run not only popular compilers, but debuggers and your favorite tools from within the editor. When shelling to DOS, VEDIT swaps itself and any desired TSRs out of memory to give you more memory than when you entered VEDIT. Only VEDIT offers you the advantages of a powerful editor without giving up the convenience of an integrated environment.

Call for your free, fully functional, evaluation copy today. See why VEDIT has been the choice of 100,000 programmers, writers and engineers since 1980.

### VEDIT Jr. - Unmatched performance for only $29.

All VEDIT editors include a pull-down menu system with "hot keys", context sensitive help, pop-up status and ASCII table, a configurable keyboard layout and flexible, unlimited keystroke macros. Perform block operations by character, line, file or column. Undo up to 1000 keystrokes - keystroke by keystroke, line by line, or deletion by deletion. Automatic indent, block indent and parentheses matching speed program development. Word wrap, paragraph formatting, justification, centering, adjustable margins and printing for word processing. Run DOS programs.

### VEDIT - A best value at only $69.

VEDIT can simultaneously edit up to 36 files and split the screen into windows. Search/replace with regular expressions. The most integrated compiler support available. Run VEDIT PLUS macros.

### VEDIT PLUS - Ultimate programmer's tool for only $185.

VEDIT PLUS adds the most powerful macro programming language of any editor. It eliminates repetitive editing tasks and permits creating your own editing functions. The macro language includes testing, branching, looping, user prompts, keyboard input, string and numeric variables, complete control over windows plus access to hardware interrupts, memory and I/O ports. Source level macro debugging with breakpoints and tracing.

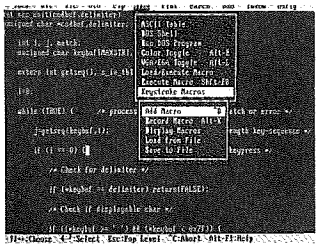
VEDIT PLUS - $185 for DOS, $285 for UNIX/XENIX. 30 Day Money-back guarantee. Discount pricing for multiple users, schools and OEMs. VEDIT Jr. site licenses start at only $250.

## VEDIT PLUS

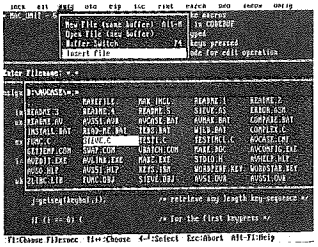
*FREE*
*Fully Functional*
*Evaluation Copy*
*Call 1-800-45-VEDIT*

- **Mouse Support**
- **Pull-down menus**
- **Columnar blocks**
- **1000 Level Undo**
- **Regular expressions**
- **Small 70K size, fast**
- **Emulate Wordstar, Word Perfect, Brief, vi, others**
- **Edit text and binary files of any size and line length**
- **Powerful macro programming language for instant "off-the-cuff" macros**

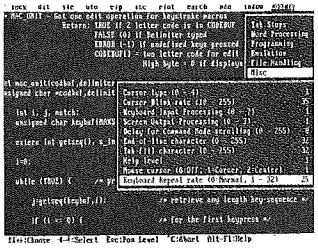
**DOS | XENIX | UNIX | QNX | FlexOS**



An intuitive user interface with pull down menus, hot keys, mouse support and context sensitive help make VEDIT easy to use, easy to learn.



Point and shoot file selection makes it a snap to edit new files, merge and split files, run macros and much more. Select files with the cursor keys or the mouse.



Just about everything about VEDIT is configurable, from the keyboard layout, to the screen colors, to the way control characters, tabs and the end of lines are displayed. Configure VEDIT with easy to use menus.

**BENCHMARKS IN 3 MEG FILE**

| | VEDIT | BRIEF | Sage | QEdit |
|---|---|---|---|---|
| Simple search | 57 sec | 1:17 min | 28 sec | Cannot |
| Save and continue | 52 sec | 3:52 min | 1:47 min | Cannot |
| Load, modify, save, exit | 21 sec | 49 sec | 1:38 min | Cannot |
| Block-column copy (40x200) | 2 sec | 30 sec | 2 sec | 2 sec |
| Delete one column in file | 9:58 min | **1:50 hour** | **1:03 hour** | Cannot |
| 60,000 replacements | 3:18 min | **1:44 hour** | **1:32 hour** | Cannot |

VEDIT is a registered trademark of Greenview Data, Inc. BRIEF is a trademark of UnderWare, Inc. Sage Profession Editor is a trademark of Sage Software Inc. QEdit is a trademark of SemWare.

* DOS version also supports Concurrent DOS, DESQview, Microsoft Windows, PC-MOS/386 and most networks. Special CRT terminal version runs over serial lines and modems.

* Free evaluation disk is fully functional and can edit small files.

# *Greenview Data*

P.O. Box 1586, Ann Arbor, MI 48106
(313) 996-1299, Fax (313) 996-1308

**Circle 133 on Inquiry Card.**

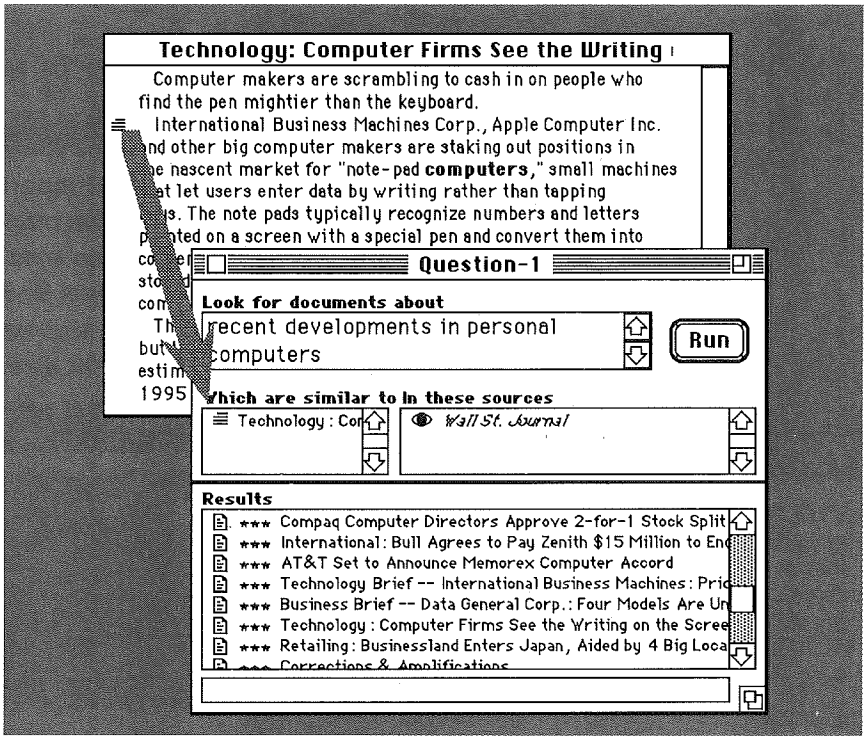



**Figure 2:** *The* `similar to` *function lets you retrieve more documents on notepad computers using relevance feedback. You then might initiate a search for additional documents with similar content. Selecting text from a section of a retrieved document helps to refine subject-matter searches or locate collateral information. You can also use the selected text to execute a new query. (Courtesy of Thinking Machines Corp.)*

With one of these tools, you can locate the position of a file and gain access to its contents. Path-driven locators search an information base for a document's name, but they do not provide a means to examine its contents.

Retrieving documents pertinent to a specific question requires *content navigation* (i.e., examining the contents of a document, or a representative abstract or index for the document, for its relevance to the question). The similarity between the question and the document's index determines a retrieval score, an indication of the likelihood that the document is pertinent.

WAISes rely on the dynamic folder to encapsulate a question. In its most passive form, it contains a question and a set of servers to target. The WAIS posts the dynamic folder to servers of known quality and functionality, and then query processing begins.

The dynamic folder executes a remote query that sends questions to the remote servers. There the questions find relevant information and return a list of document titles (document pointers) encapsulated within the originating folder to the local WAIS system. The results from the query may initially include a list of documents with fair, good, or high similarities.

Now you can refine your query strategy by perusing the document titles to determine which are the most appropriate documents. WAIS technology, in the form of the WAIStation user interface (see reference 5), assists this process through a content-associativity function known as `similar to`.

The `similar to` function informs the WAIS user interface that a document is "interesting." The server uses this information to find other documents that are similar to the one you have chosen. This search strategy, an embedded component of WAISes, represents a significant improvement over traditional database methods, such as Structured Query Language (SQL) and Boolean search.

This form of query execution is known as *relevance feedback*. It lets you extend the query to incorporate a "more-like-that-one" functionality and lets you retrieve documents that have similar contents. The WAIS user interface is organized around the English language, and English-language-oriented query structures are easier to use than SQL.

The `similar to` function is like working with a reference librarian. First, you state the topic of your research, which the librarian translates into queries. After you examine the results of the queries, you indicate which results were on the mark; thus, the librarian gains a better understanding of your needs and can improve the search.

With relevance feedback, WAISes can retrieve documents with greater ease and speed. You no longer need to alter a SQL Boolean operator to adjust the query filter; instead, you can ask for "more documents like this one."

Dynamic folders can also possess *vitality*, which gives the folder a continuous charter to execute queries periodically and update its contents with new material. A folder's charter expresses purpose, intent, and the goal that you want the query to accomplish. You can build the folder to periodically poll servers known to receive frequently updated material that matches its charter.

If the search retrieves an interesting document, WAISes let you select a portion of the text and use it as an adjunct to the initial query. Selecting text from a portion of a document that may contain some particularly topical or relevant information and using it to refine the search is an innovative approach for exploring subjects (see figure 2).

WAISes also let you chain questions by taking the results of a previous search, starting a new question with different subject matter, and dragging the previous results into the `similar to` menu box (see figure 3). Chaining questions can either broaden or narrow a search, depending on the relevance-feedback results.

The recursive capacity of dynamic folders to initiate "sibling" folders demonstrates the WAIS potential to harness and refine subject matter. Query refinement alters the charter of a dynamic folder. Sibling dynamic folders execute directed searches and can have an autonomous authority to broaden the range of server choices.

Controlling the extent of search expansion is a critical issue. For individuals, cost can be an overwhelming concern. WAIS technology does not yet contain an accounting system to govern search criteria. Participating information services will have to engineer this element of the technology themselves.

## WAIS Protocol

WAISes promote connectivity and access to remote electronic-information sources through a standard protocol, the WAIS protocol. This protocol is an extension of the National Information Standards Organization (NISO) Z39.50-1988 specification, which defines an interface to remote information-retrieval services and library-protocol applications. The Z39.50 standard is the backbone of the WAIS protocol and the foundation for WAIS applications development.

**With relevance feedback, WAISes can retrieve documents with greater ease and speed.**

Incorporating the Z39.50 standard into the WAIS protocol frees developers to build articulated user interfaces for WAIS applications. The interface standard isolates the server's text-retrieval method, such as SQL, giving the application a transparent access mode. The particulars of database queries are hidden beneath the interface. A developer only needs to be sure that the server possesses an equivalent functionality to conduct remote information-retrieval transactions from a local WAIS workstation.

Concealing the server's implementation through the WAIS protocol is important in another respect as well. Isolating the implementation implies that you can specify a single, more palatable query language. The WAIS protocol also lets you use an English-language-style query

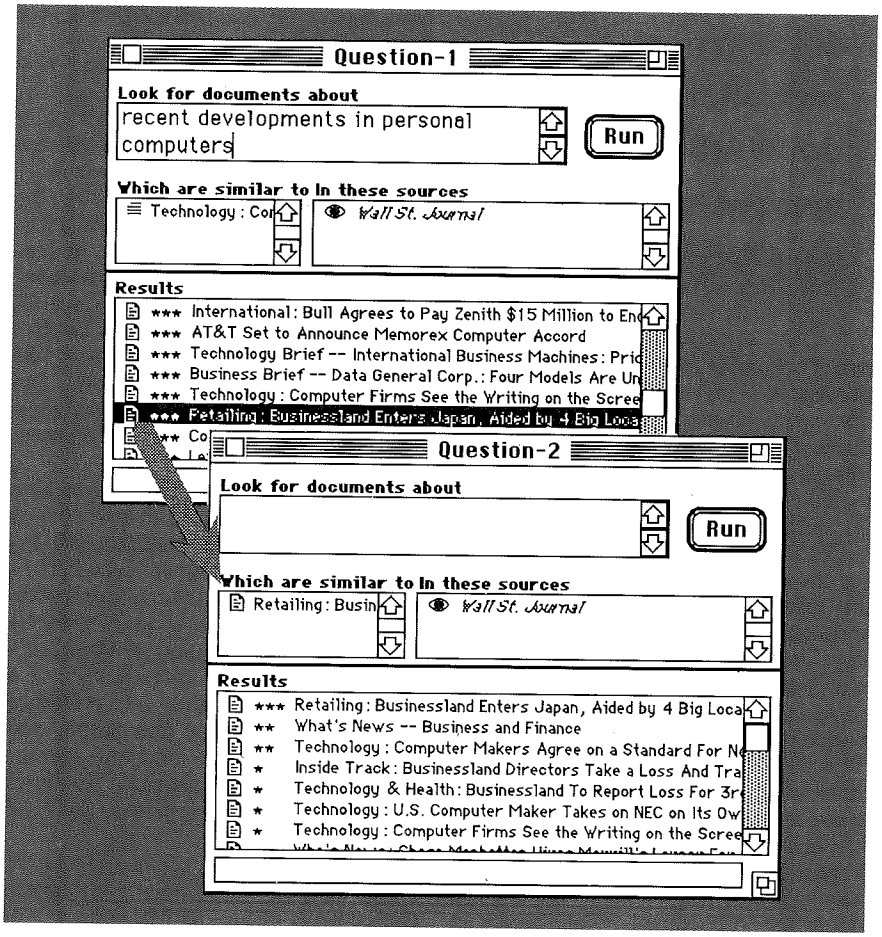



**Figure 3:** *Chaining questions permits you to use a query on multiple information sources by opening a new question and dragging previous query results into the* `similar to` *field. You can also apply the* `similar to` *operation to invoke a new document search, as in this example. (Courtesy of Thinking Machines Corp.)*

lexicon instead of cryptic SQL or fourth-generation languages. When you find a document that is appropriate, the WAIS protocol automatically handles the download process from the server. This is quite different from existing services, where manual file-capture mechanisms require vigilance. With the WAIS protocol, all documents look like they are local to your system.

The WAIS protocol incorporates two important modifications that the NISO Z39.50 standard does not address. First, it permits hypermedia document transport. Most documents today are composed primarily of ASCII text codes and sequences, but the next generation of documents, constructed from hypermedia and multimedia sources, integrates images and fully formatted text. These media forms are rapidly becoming popular and conventional.

**While still a research project that is undergoing development and refinement, the WAIS holds immense promise.**

Second, the WAIS protocol is stateless for the server. It does not have to keep any information about the client between transactions, because the user's state is kept on the local workstation. Every search or retrieval operation is a separate process. The contexts are decoupled under the statelessness of the protocol. This decoupling lets you make a search, store away the document pointer, and retrieve it later.

Further, you can use a dynamic folder to pass one of these document pointers to someone else who can also retrieve the document. A document pointer is like an International Standard Book Number for the electronic age. (The ISBN is a unique identification assigned to each publication.) Passing a document pointer conforms with copyright law and lets you easily return to the document source instead of making copies.

The WAIS protocol is designed to transport information through modems, X.25 communications, or network backbones. This flexibility provides an enormous framework within which to conduct retrieval transactions. For example, with a portable computer, you could connect with a WAIS hub through a modem and post dynamic folders, directing the query results to be routed to your office system for later examination.

## Retrieval Technology

The computing infrastructure needed to implement WAISes varies with a server's functionality. A Library of Congress WAIS, with 25 terabytes of data, could not expeditiously dispatch queries and function if a serial computer were used to process the information. For a problem of this magnitude, massive parallelism is needed. The Connection Machine's Text-Retrieval System is a viable information-retrieval system for gigabyte-size databases.

The DowQuest service from Dow Jones runs on the Connection Machine. The service incorporates approximately 1 gigabyte of original text derived from over 400 sources. The *Wall Street Journal*, the *Washington Post*, *Barron's*, *Fortune*, *Forbes*, and several regional business and technical journals are included, covering the previous eight calendar months. The search time with a 100-word query composed of typed English and relevance feedback (e.g., "more like that one") is less than half a second. The system can provide access to many gigabytes of text and to thousands of users interactively.

The projections for the Connection Machine system indicate that when it is scaled to a 1-terabyte database with 10-word queries, obtaining an answer within 10 seconds or less is highly probable. This performance is accomplished by harnessing the Connection Machine's 65,536 separate processors to execute a parallel index algorithm (see reference 6). These estimates are phenomenal and truly indicative of the computing power manifest in parallel systems. No serial machine can even come close to this level of performance.

The Connection Machine system generates these results by searching the entire contents of an archive, not a representative abstract of a keyword frequency table. Each document within the archive is used to determine a match. This is not typical for systems organized around serial computers, and it is another dramatic demonstration of parallel-computing technology.

The cost of a system like the Connection Machine runs in the millions of dollars. But a Macintosh with a 100-megabyte hard disk drive or a 386-based PC can serve the typical WAIS user.

## Immense Promise

The prototype WAIS user interface and protocol are currently being beta-tested at Thinking Machines, Apple Computer, and Dow Jones News/Retrieval. Thinking Machines, the principal developer of the WAIS architecture and software, plans to share the WAIS protocol free of charge and hopes to help user-interface developers build interfaces to WAIS servers.

While still a research project that is undergoing development and refinement, the WAIS holds immense promise. Information commerce, buoyed through the widespread acceptance of computer systems and networks, forces individuals and companies to expedite transactions and simplify activities. These coveted sources of efficiency stand out as prominent allies of competitive advantage. ■

**ACKNOWLEDGMENT**

*I'd like to thank Annie Komanecky, Franklin Davis, Ben Rewis, and Brewster Kahle of Thinking Machines for their assistance during the preparation of this article.*

**REFERENCES**

1. Hillis, D. *The Connection Machine.* Boston, MA: MIT Press, 1985.
2. Kahle, B. "Wide Area Information Server Concepts." *Thinking Machines Technical Memo DR89-1.* Cambridge, MA: Thinking Machines Corp., 1989.
3. Markoff, J. "Computer Project Would Speed Data." *New York Times*, 8 June 1990, sec. A, p. 1.
4. Markoff, J. "Creating a Giant Computer Highway." *New York Times*, 2 Sept. 1990, Part III, p. 11.
5. "WAIStation: A User Interface for Wide Area Information Servers (User Guide, Prototype Version)." Cambridge, MA: Thinking Machines Corp., 1990.
6. Stanfill, C., R. Thau, and D. Waltz. "A Parallel Indexed Algorithm for Information Retrieval." *Thinking Machines Corp. Technical Report DR 90-2.* Cambridge, MA: Thinking Machines Corp., 1990.

*Richard Marlon Stein is a software consultant and freelance writer from Van Nuys, California. He has a B.S. in physics from the University of California at Irvine. You can reach him on BIX c/o "editors."*

***Butterfly Alphabet,*** a poster showing all the letters of the English alphabet and the Arabic numerals, appeared on the cover of SCIENCE NEWS, 6/16/90. Kjell Sandved photographed the amazing forms on the wings of moths and butterflies. This spectacular full-color 18″ x 24″ poster is now available for just $10.00. Order yours today. We'll include the SCIENCE NEWS article describing the poster's origin.

ButAlphPost

*Science News Books,* 1719 N Street, NW, Washington, DC 20036

Please send me _____ copy(ies) of *Butterfly Alphabet.* I include a check payable to Science News Books for $10.00 each plus $2.00 postage and handling. (Maximum, $2.00) Domestic orders only.

Name _____

Address _____

City _____

State _____ Zip _____

Daytime Phone # (_____) _____
(used only for problems with order)

RB1440

# *Watch the World Change Before Your Eyes With the Science News Pangaea Mug!*

**All the wonders of the world's geologic history come to life on this magic mug. When cold, the mug shows the supercontinent Pangaea as it existed 200 million years ago when all of the Earth's continents existed as a single land mass. When filled with hot liquid the supercontinent breaks apart into the seven continents we recognize today. This is the mug for every coffee, tea or cocoa drinker on Earth!**


*Beige 12-ounce mug is printed in blue and brown with SCIENCE NEWS logo on reverse, $7.00 each.*

**Science News Books** PMug
1719 N Street, NW, Washington, DC 20036

Please send me _____ **Pangaea Mug(s).** I include a check payable to Science News Books for $7.00 plus $2.00 postage and handling (total $9.00) for each mug. Domestic orders only.

Name _____

Address _____

City _____

State _____ Zip _____

Daytime Phone (_____) _____
(used only for problems with order)

RB1441

# Biomedicine

## Mom's symptomless herpes threatens baby

Herpes can be deadly for newborns who acquire the viral infection from their mothers during labor or delivery. To complicate matters, only about 25 percent of adults with genital herpes display telltale symptoms (SN: 6/28/86, p.410). Seattle-based researchers have now tested a group of mothers during labor for asymptomatic but infectious herpes, and have reached some disturbing conclusions. Their data suggest that such screening will neither identify all women who risk passing the virus to their newborns, nor allow physicians to save infected babies from a devastating outcome.

Zane A. Brown and his co-workers at the University of Washington identified 56 women with active, asymptomatic herpes infections among the 15,923 laboring women tested at two local hospitals between 1984 and 1989.

Type 1 herpesvirus, the "oral" form usually associated with cold sores, rarely infects the genitalia. However, says Brown, "our data indicate that when it is present [in the genitalia], it transmits more readily [than Type 2] to infants." Three of the five women (60 percent) with Type 1 herpes infected their babies, compared with seven of the 51 women (14 percent) with active Type 2 herpes, the researchers report in the May 2 NEW ENGLAND JOURNAL OF MEDICINE.

The Type 1 infection almost never harms a newborn, Brown observes. In his study, all infants contracting Type 1 herpes developed normally. By contrast, one of the seven infants with Type 2 herpes died, and four developed disabling encephalitis.

"The really big risk of neonatal infection and damage or death occurs if a woman first acquires [Type 2] herpes late in her pregnancy," Brown says. And fully one-third of the mothers with asymptomatic herpes in this study were experiencing their first, or "primary," episode of this periodically recurring disease, he adds. Although the infected infants were identified within 24 hours of birth — far earlier than usual — and immediately treated with antiviral drugs, "we didn't significantly change the ultimate outcome," Brown says. "Kids with Type 2 disease got sick no matter what we did."

He notes another unsettling finding: Unlike most adults, infants carrying antibodies to the Type 1 virus did not contract a milder-than-normal version of the Type 2 disease.

Fortunately, the more serious, Type 2 virus "transmits reluctantly," Brown concludes. But transmission may occur in unexpected ways. He and his colleagues speculate that in eight of the babies with herpes infections, physicians may have unwittingly opened a portal for the virus by inserting electrodes into the fetus' scalp to monitor prenatal heartbeats.

## A gold-plated test for Lyme disease

Current tests for Lyme disease detect the antibodies produced when a person's immune system responds to *Borrelia burgdorferi*, the tick-borne bacteria that cause the disease. But because some people are slow to make such antibodies, the test doesn't always provide an accurate diagnosis. If left untreated, Lyme disease can cause chronic arthritic symptoms.

Scientists have now developed a prototype test that directly spots bits of *B. burgdorferi*. Working at the Rocky Mountain Laboratories of the National Institute of Allergy and Infectious Diseases in Hamilton, Mont., the group created gold-tagged antibodies that home in on two of the bacteria's surface proteins. The gold enables scientists to image bacteria-binding antibodies in the blood with an electron microscope, thereby clinching the microbes' presence, the researchers report in the June JOURNAL OF CLINICAL MICROBIOLOGY.

David W. Dorward, who directed the work, notes that most medical laboratories lack the electron microscopes needed to conduct the new test. However, he says, the gold-labeled antibodies could be adapted for use in a routine diagnostic test.

# Computers

*Ivars Peterson reports from San Jose, Calif., at the Physics Computing '91 conference*

## Navigating the information swamp

The ubiquitous lab notebook, with its dog-eared corners, stained pages and scribbled entries, may one day give way to an electronic analog that permits not only the recording of data but also the sharing of information among researchers scattered throughout the world. Researchers at Baylor College of Medicine in Houston have developed a sophisticated, computer-based scheme, called the Virtual Notebook System, that allows its user to gather, organize and annotate information selected from a variety of sources.

With such a notebook, a medical researcher interested in the diagnosis of a certain ailment, for example, can readily assemble a package consisting of X-ray images, personal comments, citations, journal articles, news items, electronic-mail extracts and other relevant pieces of information. Moreover, the researcher can instantly share that information with others who use the same system, even if they are thousands of miles away. "You can even write in someone else's notebook," says Kevin B. Long, who directed the project.

Designed to facilitate collaboration, the system's key element consists of software that masks the underlying maze of computers and computer networks that often stands in the way of efficient and convenient communication among researchers working with different computer equipment. The Virtual Notebook System also incorporates a new programming approach for simplifying the indexing and retrieval of information stored in computers. A specially programmed, information-seeking computer — known as the Wide Area Information Server and developed under the direction of Brewster Kahle of Thinking Machines Corp. in Cambridge, Mass. — responds to requests typed in English. Users don't have to know exactly how to find the information they need; nor do they have to remember any special instructions to locate data.

Best suited for groups of researchers already linked by computer networks, the Virtual Notebook System may prove a crucial component of large collaborative efforts. Officials with the Superconducting Super Collider are investigating the system as a possible means of sharing and analyzing experimental data when the accelerator is eventually completed.

## Exploring the virtual wind

Calculations of the direction and speed at which air flows past a complicated, three-dimensional object, such as an airplane, generate huge quantities of data. Conventional two-dimensional graphic images derived from these data often fail to convey the flow's complexity. Now, a team of researchers has assembled a primitive, prototype system for exploring such flow patterns, in effect allowing an investigator to step into and interact with a computer-generated environment.

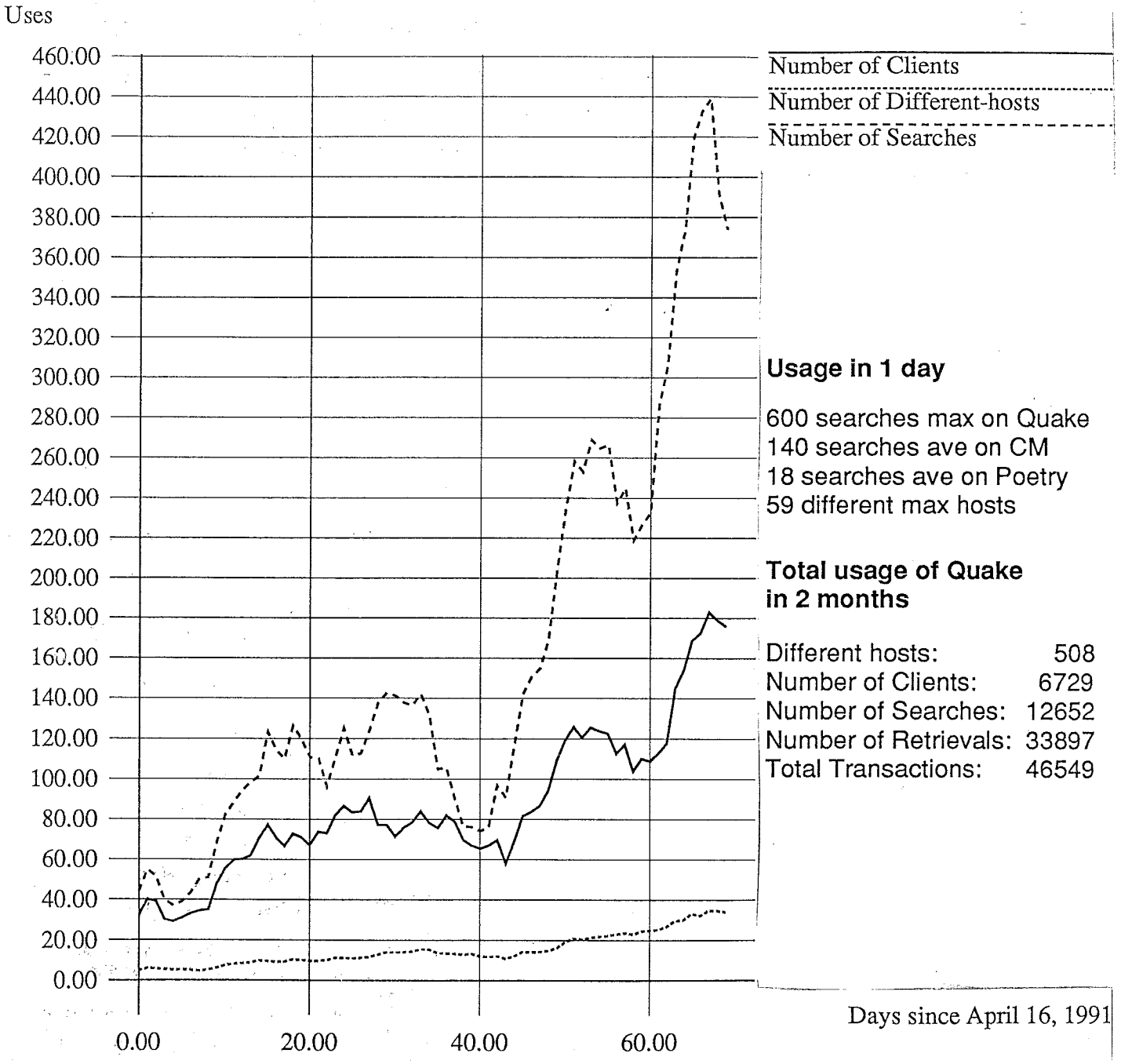
Steve Bryson and Creon Levit of the NASA Ames Research Center in Mountain View, Calif., used commercially available components to create their prototype system, known as the Virtual Windtunnel.

Through computer graphics and special input devices, the system creates the illusion of being surrounded by a flow. The user looks through a boom-mounted device resembling a diver's mask, which contains two small television sets to produce a wide-angle, stereoscopic image. A computer tracks the viewer's head position and generates the appropriate views. The user also wears a flexible glove fitted with sensors to manipulate the image in various ways. For example, to visualize the direction of flow in a particular region, a researcher can use the glove to specify the starting point for a computer-rendered stream of smoke, and then walk around to see the resulting flow pattern from different angles.

"It puts the computations right in front of the researcher," Bryson says. "It allows real interaction with the data."

# WAIS

## WAIS Daily Usages on Quake.Think.Com




**Usage in 1 day**

600 searches max on Quake
140 searches ave on CM
18 searches ave on Poetry
59 different max hosts

**Total usage of Quake in 2 months**

| | |
|---|---|
| Different hosts: | 508 |
| Number of Clients: | 6729 |
| Number of Searches: | 12652 |
| Number of Retrievals: | 33897 |
| Total Transactions: | 46549 |

**Countries Using WAIS:**

Austria, Canada, Denmark, Finland, France, Germany, Holland, Italy, Mexico, Norway, Sweden, Switzerland, USA

**Thinking Machines Corporation**

# Biomedicine

## Mom's symptomless herpes threatens baby

Herpes can be deadly for newborns who acquire the viral infection from their mothers during labor or delivery. To complicate matters, only about 25 percent of adults with genital herpes display telltale symptoms (SN: 6/28/86, p.410). Seattle-based researchers have now tested a group of mothers during labor for asymptomatic but infectious herpes, and have reached some disturbing conclusions. Their data suggest that such screening will neither identify all women who risk passing the virus to their newborns, nor allow physicians to save infected babies from a devastating outcome.

Zane A. Brown and his co-workers at the University of Washington identified 56 women with active, asymptomatic herpes infections among the 15,923 laboring women tested at two local hospitals between 1984 and 1989.

Type 1 herpesvirus, the "oral" form usually associated with cold sores, rarely infects the genitalia. However, says Brown, "our data indicate that when it is present [in the genitalia], it transmits more readily [than Type 2] to infants." Three of the five women (60 percent) with Type 1 herpes infected their babies, compared with seven of the 51 women (14 percent) with active Type 2 herpes, the researchers report in the May 2 NEW ENGLAND JOURNAL OF MEDICINE.

The Type 1 infection almost never harms a newborn, Brown observes. In his study, all infants contracting Type 1 herpes developed normally. By contrast, one of the seven infants with Type 2 herpes died, and four developed disabling encephalitis.

"The really big risk of neonatal infection and damage or death occurs if a woman first acquires [Type 2] herpes late in her pregnancy," Brown says. And fully one-third of the mothers with asymptomatic herpes in this study were experiencing their first, or "primary," episode of this periodically recurring disease, he adds. Although the infected infants were identified within 24 hours of birth — far earlier than usual — and immediately treated with antiviral drugs, "we didn't significantly change the ultimate outcome," Brown says. "Kids with Type 2 disease got sick no matter what we did."

He notes another unsettling finding: Unlike most adults, infants carrying antibodies to the Type 1 virus did not contract a milder-than-normal version of the Type 2 disease.

Fortunately, the more serious, Type 2 virus "transmits reluctantly," Brown concludes. But transmission may occur in unexpected ways. He and his colleagues speculate that in eight of the babies with herpes infections, physicians may have unwittingly opened a portal for the virus by inserting electrodes into the fetus' scalp to monitor prenatal heartbeats.

## A gold-plated test for Lyme disease

Current tests for Lyme disease detect the antibodies produced when a person's immune system responds to *Borrelia burgdorferi*, the tick-borne bacteria that cause the disease. But because some people are slow to make such antibodies, the test doesn't always provide an accurate diagnosis. If left untreated, Lyme disease can cause chronic arthritic symptoms.

Scientists have now developed a prototype test that directly spots bits of *B. burgdorferi*. Working at the Rocky Mountain Laboratories of the National Institute of Allergy and Infectious Diseases in Hamilton, Mont., the group created gold-tagged antibodies that home in on two of the bacteria's surface proteins. The gold enables scientists to image bacteria-binding antibodies in the blood with an electron microscope, thereby clinching the microbes' presence, the researchers report in the June JOURNAL OF CLINICAL MICROBIOLOGY.

David W. Dorward, who directed the work, notes that most medical laboratories lack the electron microscopes needed to conduct the new test. However, he says, the gold-labeled antibodies could be adapted for use in a routine diagnostic test.

# Computers

*Ivars Peterson reports from San Jose, Calif., at the Physics Computing '91 conference*

## Navigating the information swamp

The ubiquitous lab notebook, with its dog-eared corners, stained pages and scribbled entries, may one day give way to an electronic analog that permits not only the recording of data but also the sharing of information among researchers scattered throughout the world. Researchers at Baylor College of Medicine in Houston have developed a sophisticated, computer-based scheme, called the Virtual Notebook System, that allows its user to gather, organize and annotate information selected from a variety of sources.

With such a notebook, a medical researcher interested in the diagnosis of a certain ailment, for example, can readily assemble a package consisting of X-ray images, personal comments, citations, journal articles, news items, electronic-mail extracts and other relevant pieces of information. Moreover, the researcher can instantly share that information with others who use the same system, even if they are thousands of miles away. "You can even write in someone else's notebook," says Kevin B. Long, who directed the project.

Designed to facilitate collaboration, the system's key element consists of software that masks the underlying maze of computers and computer networks that often stands in the way of efficient and convenient communication among researchers working with different computer equipment. The Virtual Notebook System also incorporates a new programming approach for simplifying the indexing and retrieval of information stored in computers. A specially programmed, information-seeking computer — known as the Wide Area Information Server and developed under the direction of Brewster Kahle of Thinking Machines Corp. in Cambridge, Mass. — responds to requests typed in English. Users don't have to know exactly how to find the information they need; nor do they have to remember any special instructions to locate data.

Best suited for groups of researchers already linked by computer networks, the Virtual Notebook System may prove a crucial component of large collaborative efforts. Officials with the Superconducting Super Collider are investigating the system as a possible means of sharing and analyzing experimental data when the accelerator is eventually completed.

## Exploring the virtual wind

Calculations of the direction and speed at which air flows past a complicated, three-dimensional object, such as an airplane, generate huge quantities of data. Conventional two-dimensional graphic images derived from these data often fail to convey the flow's complexity. Now, a team of researchers has assembled a primitive, prototype system for exploring such flow patterns, in effect allowing an investigator to step into and interact with a computer-generated environment.

Steve Bryson and Creon Levit of the NASA Ames Research Center in Mountain View, Calif., used commercially available components to create their prototype system, known as the Virtual Windtunnel.

Through computer graphics and special input devices, the system creates the illusion of being surrounded by a flow. The user looks through a boom-mounted device resembling a diver's mask, which contains two small television sets to produce a wide-angle, stereoscopic image. A computer tracks the viewer's head position and generates the appropriate views. The user also wears a flexible glove fitted with sensors to manipulate the image in various ways. For example, to visualize the direction of flow in a particular region, a researcher can use the glove to specify the starting point for a computer-rendered stream of smoke, and then walk around to see the resulting flow pattern from different angles.

"It puts the computations right in front of the researcher," Bryson says. "It allows real interaction with the data."

MACWEEK
THE NEWSWEEKLY FOR MACINTOSH MANAGERS
05.14.91 Volume 5 Number 19

# WAIS promises easy text retrieval

## *Prototype links Mac, Connection Machine*

By Henry Norr

Cupertino, Calif. — Thinking Machines Corp., a pioneer in the development of high-powered parallel-processing supercomputers, has joined with Apple, Dow Jones & Co. and KPMG Peat Marwick to develop a new technology designed to simplify the retrieval of textual information stored in personal files, corporate records and remote databases.

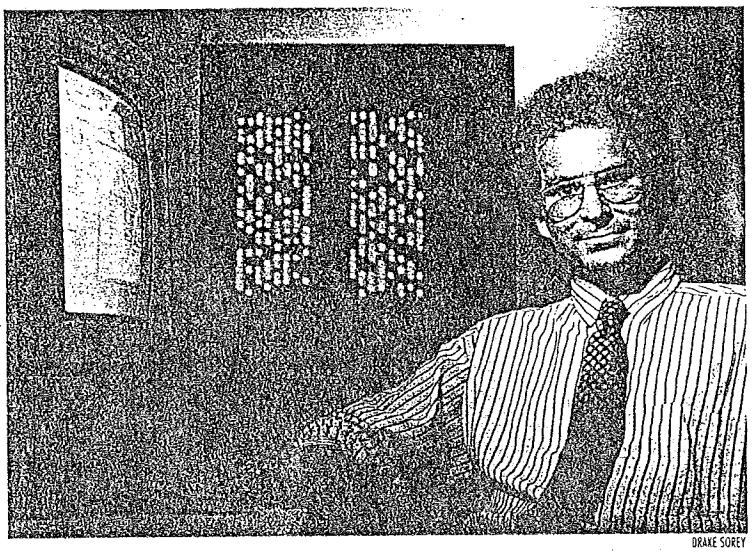
Called the Wide Area Information Server (WAIS) project, the collaborative venture has been under way for almost two years. Peat Marwick recently completed a four-month experiment with the system, using WAIStation, a prototype Mac front end developed by Thinking Machines of Cambridge, Mass. Engineers from Apple's Advanced Technology Group have combined the WAIS technology with a custom interface to build a prototype personal electronic newspaper.

The WAIS project was designed in part to address problems caused by the proliferation of electronic data within large organizations.

"Corporations are starting to gag on gigabytes of word processing files, memos, reports, articles and E-mail archives," said Brewster Kahle, WAIS project leader for Thinking Machines. "Corporate memory is stored in this form, but executives have no easy way to get at it."

But the WAIS project was intended from the beginning, Kahle said, to be more than a traditional executive information system working only within corporate bounds. The objective was to lay the foundations for a scalable system that would allow users to tap a variety of data sources, including large commercial databases, through a uniform interface. Users, according to the plan, should be able to search for any available information without having to master the internal organization and query techniques of each source.

*See Thinking Machines, Page 24*



DRAKE SOREY

**Brewster Kahle, WAIS project leader, helped develop an experimental text-retrieval system that can use a Thinking Machines supercomputer as a server and the Mac as a front end.**

# Peat Marwick tries 'partner-friendly' system

When officials of Thinking Machines Corp., Dow Jones & Co. and Apple first broached the concept of the Wide Area Information Server (WAIS) with KPMG Peat Marwick, representatives of the accounting giant were intrigued but cautious, according to Brewster Kahle, project leader for Thinking Machines.

They weren't interested, he said, in another complex querying application that busy tax consultants, accountants and managers would never bother to use. But they agreed to participate in the project, according to Kahle, on the promise of a system that would be genuinely "partner-friendly," with "no algebra — no ifs, ands or buts."

After a year of preliminary work, an experimental WAIS R&D project went on-line at Peat Marwick last October. About 10 users at the company's Montvale, N.J., headquarters, including "very senior partners," took part in the experiment, along with two others in Manhattan and 10 more on the West Coast, according to Robin Palmer, senior manager and WAIS project leader at KPMG Peat Marwick in San Jose, Calif. The remote users were connected by leased lines to a WAIS server running on a Connection Machine, a Thinking Machines parallel-processing system, installed in Montvale.

The Peat Marwick experiment relied on WAIStation, a Mac-based client software program developed by Thinking Machines, as a front end. To prepare a query, users need only enter the subject they are interested in, in English — "IBM and Motorola," for instance, or "recent developments in personal computers" — in a text field labeled "Look for documents about." They then drag icons representing possible sources, local or remote, into another field.

When the query is run, the Macintosh-based front end encodes the search string according to the WAIS protocol and passes it to the specified servers. Each server translates the query into its own language, locates matching articles and returns the results to the front end.

The WAIStation application then displays headlines for each article; the citations are ranked according to probable relevance, based on algorithms that consider the position, frequency and proximity of desired terms within the text.

By double-clicking on the headline, users can get the full text of any of the articles. And if the user drags the most useful titles into a bin labeled "Similar to" and reruns the search, the system will track down additional articles that share a large number of words with those selected.

Peat Marwick completed its WAIStation testing in February. In part because the cost of maintaining a real-time wide-area link among its many offices would be "substantial," according to Palmer, the company has not made a commitment to the system and is still considering a variety of alternatives. But, he said, "we are still extremely interested in the WAIS concepts. It's a most promising technology." — *By Henry Norr*

## Thinking Machines

*From Page 22*

The WAIS system has three components:

▸ **Server software.** Any information source capable of locating and presenting text in response to a request in WAIS format can function as a server; the source can be on the user's own machine, on a LAN or at a remote site connected by modem. The WAIS client software can keep track of multiple servers, search any or all in response to a single request and consolidate the results.

Thinking Machines now includes the WAIS text-indexing and retrieval software free with its Connection Machines, a line of massively parallel systems that range in price from $100,000 to $5 million, according to Kahle. In addition, the companies participating in the project developed a sample server that runs on standard Unix systems. But any text-retrieval program on any platform, including the Mac, could be adapted to function as a WAIS server.

▸ **Protocol.** To foster the development of WAIS-compatible data sources, the four companies created an open protocol for transmitting queries and responses. It is based on an existing standard, the National Information Standards Organization's Z39.50 protocol, but is enhanced in several ways, such as by the addition of support for audio and video information.

▸ **Clients.** WAIS was designed to support a variety of interfaces running on various platforms and tailored to different niches.

The system does not rely on a specialized query language; the front end simply passes English-language search strings entered by the user to the server.

In addition to the prototype WAIStation interface and Apple's experimental personal newspaper, front ends already are available for the X Window System and GNU emacs, an extensible text editor that runs under a freely distributed Unix-like operating system developed at the Massachusetts Institute of Technology in Cambridge.

To promote the WAIS concept, Thinking Machines is making source code for the system available over the Internet or by mail. The code comes free of charge but without support. Using the software, programmers at MIT and elsewhere already have created more than 20 WAIS servers, including a poetry server, a weather server and a catalog of government programs. Thinking Machines will maintain a publicly accessible directory of servers, which will include descriptions of all known servers and special files that allow WAIS front ends to plug into them.

Business Day

The New York Times

# For Shakespeare, Just Log On

## Large PC Libraries Are Being Developed

By JOHN MARKOFF

The development of a nationwide data network will allow personal computer users to tap sources as large as the Library of Congress or receive their own personalized electronic newspapers.

Several innovations, taken together, have already demonstrated that searching vast computer data bases can be easier than consulting a card catalogue, and not nearly as difficult or expensive as computer searches are today. Computer users might read some Dickens more readily than they could check out David Copperfield from the local library.

Those in the industry say that users with little computer skills will soon be able to search through several terabytes of information, or several trillion characters of text, in seconds. The Library of Congress, with 80 million items, contains an estimated 25 terabytes of information.

Already, an experimental computer library has linked 150 universities to 40 sources of information, ranging from National Institutes of Health data to corporate documents and Shakespeare's plays. New software allows users to browse or zero in on particular information.

As methods of retrieving information are standardized and perfected, industry executives and computer scientists say, thousands of new services, ranging from electronic newspapers to the computer equivalent of free public libraries, will blossom. "Everyone is realizing how important it is to get into the mass market for information," said Thomas Koulopoulos, president of Delphi Consulting Group, a Boston market research firm.

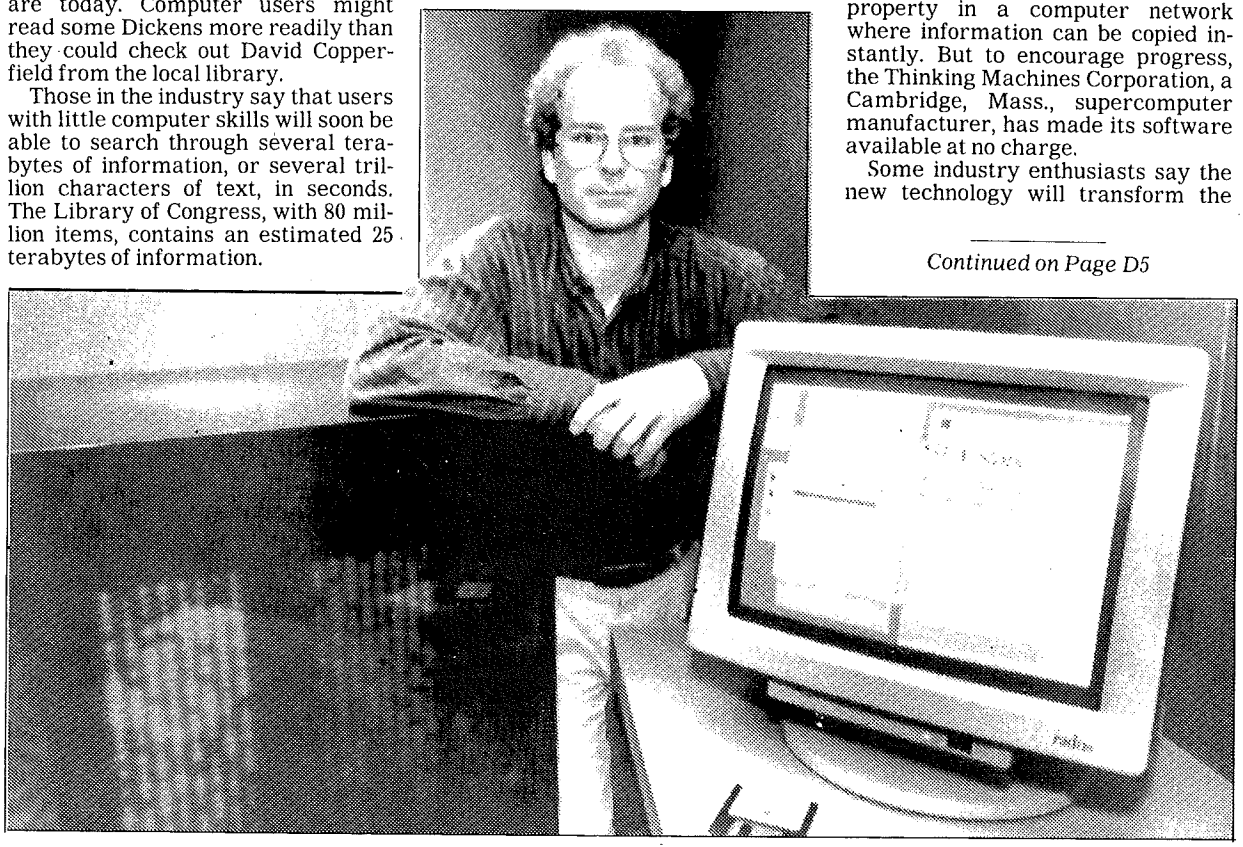
Such ready access to huge amounts of computerized information has been the dream of many in the industry. But a lack of computing power, effective software and high-speed digital networks has stalled progess until recently.

If many of the technical problems are being solved, major business and political disputes remain. The researchers acknowledge that they must resolve several questions of privacy and pricing before they can put the new methods to commercial use.

Many sources of information, like government documents, might be available free, but other services, including electronic newspapers, will be available only to those who pay. The industry has yet to settle on ways to protect and charge for intellectual property in a computer network where information can be copied instantly. But to encourage progress, the Thinking Machines Corporation, a Cambridge, Mass., supercomputer manufacturer, has made its software available at no charge.

Some industry enthusiasts say the new technology will transform the

*Continued on Page D5*



Mike Theiler for The New York Times

Brewster Kahle was the leader of the development team at the Thinking Machines Corporation for a nationwide computerized library system. His team's software links a CM2A Connection Machine, left, with a personal computer or work station like the Apple Macintosh II at right. Using high-speed data highways, the two machines can function together although they may be thousands of miles apart.

# For Shakespeare, Just Log On

*Continued From First Business Page*

way computerized information is sold. Mitchell Kapor, the founder of the Lotus Development Corporation, predicts the growth of a new industry as significant as the personal computer business. Some companies, like Dow Jones & Company, that already provide computerized information over telephone lines have taken part in developing the new computer library.

**The Search Is Simplified**

In 1989, Thinking Machines enlisted the support of Dow Jones, Apple Computer Inc. and the KPMG Peat Marwick accounting and consulting firm to design the computer library, called Wide Area Information Servers, or WAIS (pronounced ways). The system permits computer users to quickly search through a huge volume of information even if it is stored at several distant locations.

The system lets users conduct searches by typing common English phrases instead of more complicated computer commands. While current systems like Dialog and Nexis require users to specify precisely the information they want, the new system can respond to a user's inferences. It initially presents a sample list of documents. The user chooses one or several, and then a "relevance feedback" program presents other documents most like the ones selected.

"This solves the problem of how to get to the information you need, getting not too much and not too little," said Esther Dyson, editor of Release 1.0, a computer industry newsletter.

> It will soon be possible to search through millions of items in seconds.

This is a sharp contrast to the way services operate today, Ms. Dyson said. A computer user may need to call seven or eight separate data bases depending on the kind of information needed.

The WAIS system lets users of Apple personal computers harness a network of Thinking Machines supercomputers and smaller "server" computers to search data bases stored by Dow Jones, KPMG and several corporations and universities. Users can also read electronic mail, enter their corporate electronic libraries and summon up a wide variety of documents, newspapers and magazines.

**A 'Corporate Memory'**

At Thinking Machines, the WAIS system serves as a "corporate memory," allowing employees to retrieve memos, documents and other internal information. Employees who may not be working together can share expertise.

"If someone did something in Los Angeles and I'm sitting in San Francisco, I may not know about the work," said Robin Palmer, a senior manager at Peat Marwick.

WAIS delivers information over the Internet, a collection of 2,600 high-speed public and private computer networks. This Government-sponsored system of data highways is rapidly being improved and turned to commercial uses.

The market for software that allows the rapid retrieval of computerized text is small but growing, according to industry analysts. In 1989, the United States had fewer than 60,000 users; by the next year, total sales were about $120 million. The Delphi Consulting Group expects the market to grow to 160,000 users and $235 million by 1992.

"Information retrieval technology is starting to spread from supercomputers all the way down to personal computers," said Brewster Kahle, a Thinking Machines scientist who has led the WAIS experiment.

The WAIS system is built on a procedure for retrieving information developed by librarians who initially set out to computerize their card catalogues. The procedure — known in the field as Z39.50 — now has the support of the Library of Congress, Apple, Sun Microsytems Inc., Next Inc., Dow Jones and Mead Data Central.

In the future, a special directory or "white pages" will keep an up-to-date list of all the separate sources on the network.

Apple has its own electronic library project, borrowing its name, Rosebud, from the movie "Citizen Kane." The three-year-old project is based on the WAIS system, but adds features including the ability for a user to develop a personalized electronic newspaper.

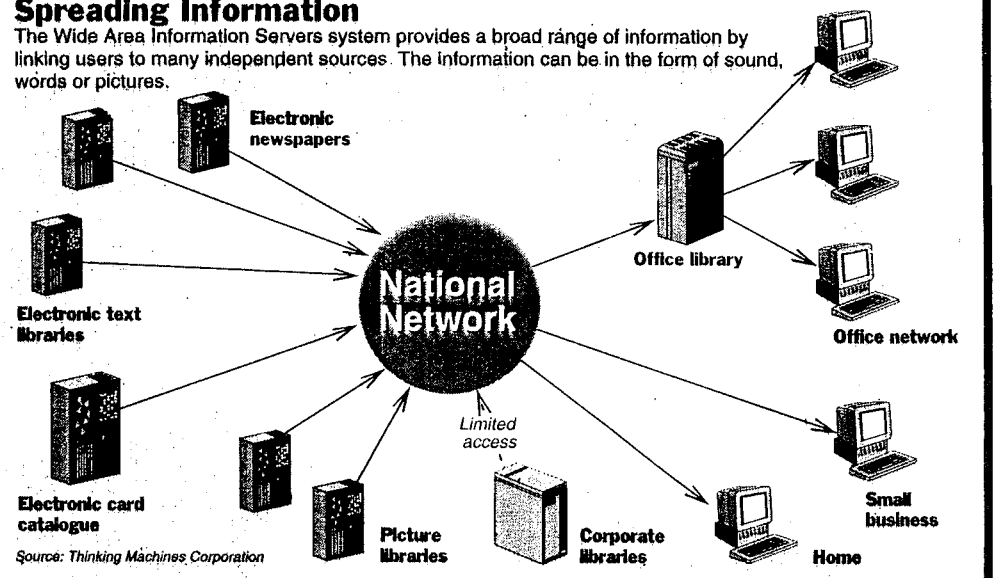
Rosebud uses special programs — called "reporters" — that let customers specify the kinds of information and news they want to retrieve from the WAIS system every day. Researchers at Apple's Advanced Technology Group said that in the future the necessary retrieval software might be a standard part of a computer's operating system.

They expect improvements in the Internet computer network to greatly lower the cost of information searches, promoting the introduction of many new services. The Government proposes to expand and improve Internet by financing a National Research and Education Network, or NREN, that could extend a high-speed computer links into schools and communities across the country.

"With things like NREN, everthing could change overnight," said Tim Oren, an Apple researcher.

## Spreading Information

The Wide Area Information Servers system provides a broad range of information by linking users to many independent sources. The information can be in the form of sound, words or pictures.



Herald INTERNATIONAL Tribune.
BUSINESS/FINANCE
** THURSDAY, JULY 4, 1991

# Having It All: Vast Data Networks Near

By John Markoff
*New York Times Service*

NEW YORK — The development of a U.S. data network will allow users of personal computers to tap sources as large as the Library of Congress or receive their own personalized newspapers.

Several innovations have already demonstrated that searching vast data bases can be easier than consulting a card catalogue, and not nearly as difficult or expensive as computer searches today.

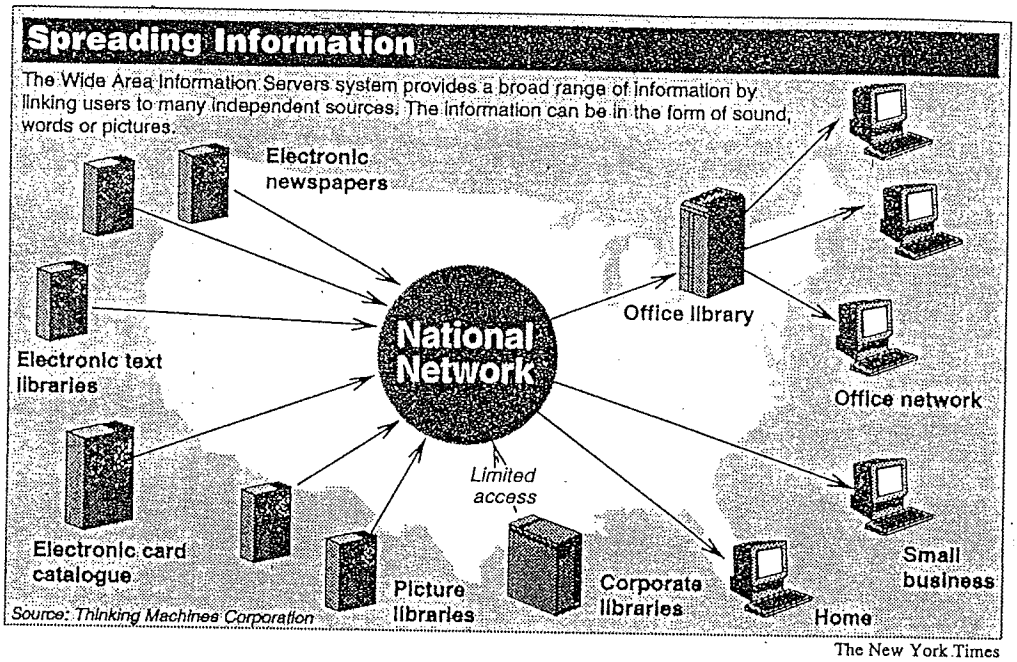
Users with minimal computer skills would soon be able to search through several terabytes of information — several trillion characters of text — in seconds. The Library of Congress, with 80 million items, contains an estimated 25 terabytes of information.

Already an experimental computer library has linked 150 universities to 40 sources of information, ranging from National Institutes of Health data to corporate documents and Shakespeare's plays.

New software allows users to browse or zero in.

As methods of retrieving information are standardized and perfected, industry executives and computer scientists say, thousands of new services, ranging from electronic newspapers to the computer equivalent of free public libraries, will blossom.

"Everyone is realizing how important it is to get into the mass market for information," said Thomas Koulopoulos, president of Delphi Consulting Group, a Boston market-research firm.



The New York Times

Such ready access to huge amounts of computerized information has been the dream of many, but a lack of computing power, effective software and high-speed digital networks stalled progress.

If many of the technical problems are being solved, major business and political disputes remain.

The industry has yet to find ways to protect and charge for intellectual property in a computer network.

To encourage progress, Thinking Machines Corp., a Cambridge, Massachusetts, computer manufacturer, has made its software free.

Some companies, like Dow Jones Corp., that already provide computerized information over telephone lines, have taken part in developing the new computer library.

In 1989, Thinking Machines enlisted the support of Dow Jones, Apple Computer Inc. and the KPMG Peat Marwick accounting and consulting firm to design the computer library, called Wide Area Information Servers, or WAIS.

The system permits computer users to quickly search a huge volume of information even if it is stored at several distant locations.

While current systems like Dialog and Nexis require users to specify precisely, the new system can respond to inferences. It presents a sample list of documents. The user chooses one or several, and a feedback program presents other documents most like the ones selected.

"This solves the problem of how to get to the information you need, getting not too much and not too little," said Esther Dyson, editor of Release 1.0, a computer industry newsletter.

A computer user may need to

See DATA, Page 16

## DATA: *Information Networks in the U.S. Are Expanding by Terabytes*

(Continued from first finance page)

call seven or eight separate data bases depending on the information needed.

The WAIS system lets users of Apple personal computers harness a network of Thinking Machines supercomputers and smaller computers to search data bases stored by Dow Jones, KPMG and several corporations and universities.

Users can also read electronic mail, enter their corporate electronic libraries and summon up a wide variety of documents, newspapers and magazines.

At Thinking Machines, the WAIS system serves as a corporate memory, allowing employees to retrieve memos, documents and other internal information.

In 1989, the United States had fewer than 60,000 users in the market for software that allows the rapid retrieval of computerized text. By the next year, total sales were about $120 million. The Delphi Consulting Group expects the market to grow to 160,000 users and $235 million by 1992.

Apple has its own electronic-library project, borrowing its name, Rosebud, from the movie "Citizen Kane." The project is based on WAIS, but adds features including the ability for a user to develop a personalized electronic newspaper.

Rosebud uses special programs, called reporters, that allow customers to specify the information and news they want to retrieve from WAIS daily.

"Information retrieval technology is starting to spread from supercomputers all the way down to personal computers," said Brewster Kahle, a Thinking Machines scientist.